联合国粮食及农业组织 | Food and Agriculture Organization of the United Nations | Organisation des Nations Unies pour l'alimentation et l'agriculture | Продовольственная и сельскохозяйственная организация Объединенных Наций | Organización de las Naciones Unidas para la Alimentación y la Agricultura | منظمة الأغذية والزراعة للأمم المتحدة

COORDINATING WORKING PARTY ON FISHERY STATISTICS

Intersessional Meetings of Aquaculture and Fisheries Subject Groups
Joint Session

Eighth Meeting of the Aquaculture Subject Group (AS)
Thirty meeting of the Fisheries Subject Group (FS)

28- 30 June 2023

Meeting place: Hybrid - NEAFC headquarters
London, United Kingdom/WebEx

Digital resources & data exchange format technical specifications in support to CWP reference data harmonization implementation guidelines (version 1.0)

Author: *ad hoc* Task Group on reference harmonization (TG-RH)

Executive Summary

The present document introduces the digital guidelines in support to the conceptual CWP RH implementation guidelines provided so far. Such digital guidelines are compound by 1) a set of online digital resources and 2) a technical specification of data exchange formats to implement fisheries data exchange reference harmonization case studies. The digital resources include *terms*, *codelists*, *mappings* and *formats* made available in a broader collaborative and working environment with focus on fisheries data interoperability named *fdiwg* (standing for *Fisheries Data Interoperability ad hoc working group*). Such set of digital asset can be used as working catalogue for CWP established standard classifications, ongoing standardization efforts and extended to other standards needed for the implementation of CWP RH case studies, but not directly covered by CWP. The data exchange formats specifications provide general format specifications to ensure proper fisheries data exchange and interoperability, as well as two data structures 1) one *generic* providing a measurement-agnostic tabular data structure, 2) one *simplified* provided a measurement-dependent structure aims to facilitate data entry in some use cases (e.g., National data clerks for regional DCRFs).

Contents

1. Background

The first version (1.0) of the CWP Reference harmonization implementation guidelines set up the framework for implementing fisheries statistical concepts defined and endorsed by CWP through data structures, with a first insight on best practices. The revised version (1.1) of the guidelines submitted for the present CWP IS further builds on the feedback and experiences collected from CWP RH case studies. While keeping a conceptual framework? through the main implementation guidelines (CWP-IS_2023/5.4.a), the present extension has been written to go deeper into digital implementation considerations, covering more technical aspects of the Reference Harmonization implementation to interoperate with reference data resources, and give technical specifications of fisheries data exchange formats to recommend when dealing with CWP RH case studies.

2. Scope and objectives

The present document extends the revised CWP Reference Harmonization implementation guidelines by providing to CWP parties digital technical specifications and guidelines to implement fisheries data exchange.

3. Contributors

- **FAO**
  - Emmanuel Blondel –NFISI, Geographic Information Systems & R Expert ; CWP TG-Reference Harmonization and TG-GeoSpatial task leader
  - Aymen Charef – FAO, OCS, Fishery Statistician; CWP Reference Harmonization task former leader
  - Marc Taconet – FAO, NFISI Team leader
  - Yann LAURENT – FAO, NFISI, Statistics and Information System Expert, WECAFC-FIRMS project manager
  - Aureliano Gentile – FAO, NFISI, Information manager
  - Anton Ellenbroek – FAO, NFISI, Information analyst
  - Alexandre Bennici – FAO, NFISS, Statistician / R developer
  - Johanna Herfault – FAO, NFISI, Fisheries specialist
- **IOTC**
  - Fabio Fiorellato – IOTC Secretariat, NFIT, Fishery Officer (Data coordinator, FIRMS Tuna Atlas task leader)
  - Emmanuel Chassot – IOTC Secretariat, NFIT, Fishery Officer
- **IRD** (Collaborating partner for the FIRMS Global Tuna Atlas)
  - Julien Barde – IRD

The document has been consolidated based on experience and feedback from CWP RH pilot use cases including:

- **Global use cases**
  - the FIRMS Global Tuna Atlas[1],
  - the NFISS statistical yearbook
- **Regional data collection frameworks**

[1] https://www.fao.org/fishery/en/collection/firms-tuna-atlas

  - the WECAFC-FIRMS regional database (known as WECAFIS - Western Central Atlantic Fisheries Information System),
  - the RECOFI MDR
- **National Fisheries Statistics and Management Information Systems**
  - the FAO Calipseo[2] platform

Additional inputs were contributed as feedback from the design of the FIRMS dcf-shiny web application, developed under FIRMS and the EC Blue-Cloud project, as solution to support implementation of data calls in response to Data collection frameworks and underlying operation of fisheries management databases.

4. Audience

The present standard technical specification targets data managers of organizations in charge of data sharing to support the implementation of the CWP standard on reference harmonization.

The specifications are of particular interest for use cases with the need to implement data calls as part of established or ad hoc Data Collection Frameworks, for example the WECAFC Data Collection Reference Framework (DCRF)[3], or the FIRMS Global Tuna Atlas yearly data calls. For use cases not directly related to data calls (e.g., FAO NFISS statistical yearbook), the specifications provide digital implementation guidelines and best practices to access and use CWP reference data sets such as codelists and mappings.

5. Components

5.1 Overview

The present standard technical specifications constitute a whole *package* made of various components including:

- **digital standardized resources**, available on the web for direct access and use by CWP RH use cases; including:
  - a first dictionary of terms usable for fishery concepts and their definitions
  - a repository of standardized code lists
  - a repository of standardized code list mappings
  - a repository digital formats specifications
- a **standard fisheries data exchange format specification** and its variants in support of data calls, and based on the above digital standardized resources

The whole package aims to be versioned and published through the use of Digital Object Identifiers (DOI) for better referencing in implementing use cases. Upon feedback from parties through the CWP lifecycle, the present standard could be revised, modified, corrected and extended to fill eventual gaps and cover additional needs (as for any standard).

[2] https://www.fao.org/fishery/en/statistics/software/calipseo
[3] https://www.fao.org/wecafc/data/dcrf

5.2 Digital resources

5.2.1 Rationale

In continuity with the work undertaken since CWP 26 to create a CWP catalogue, and a general need to interoperate CWP RH case studies with reference data material, digital resources have been progressively made available.

The options and proof-of-concepts initially highlighted in the first version of the CWP RH implementation guidelines (usage of CKAN, EBX5 tool or MS Excel) have proven to show more constraints than assets for their exploitation by case studies, either due to lack of standardization, interoperability, simplicity, and even accessibility (tool requiring authorization).

An open, collaborative, pragmatic, and simple solution was then needed, at least as working environment, to enable CWP reference data streams in support of implementation case studies. For these reasons, a working environment was set up with the open Github platform, with a series of repositories, aligning with practices already in place for case studies (e.g., Global Record of Stocks & Fisheries – GRSF at https://github.com/GRSF ; or the FIRMS Global Tuna Atlas – GTA at https://github.com/firms-gta).

Digital resources may target resources already stamped as CWP standards including well-established ones such as ASFIS, ISSCFG; new standards such as the *fishing fleet* (used in the FIRMS Global Tuna Atlas, and endorsed by CWP 27), new code lists to be drafted as part of CWP TG activities (e.g., catch types, effort units), but also ad hoc reference data material not yet proposed nor endorsed by CWP but needed to operate CWP RH case studies (e.g., measurement?), as well as other fishery non-specific standards (e.g., fishing mode?).

For this reason, a working name without mention of CWP was used for the base repository of digital resources, i.e., *fdiwg* – standing for *Fisheries Data Interoperability ad hoc working group*.This name is not new to CWP as it was already introduced to CWP in the past[4] (Copenhagen, 2017) as initiative carried out with the EC Blue-Bridge project to create a broader ad hoc working group under the Research Data Alliance[5]. The Github repository was created as a follow-up of this initiative.

The open repository approach aims to foster collaboration from CWP parties and beyond (e.g., collaborating partners identified the FIRMS partnership), and hence, to foster the participation of CWP parties to CWP Reference harmonization task group activities, in relation with other TGs (e.g., TG-Catch, TG-Effort).

Digital resources were made available under the overall concern to implement FAIR (Findable Accessible Interoperable and Reusable) data principles[6], and to follow a pragmatic cataloguing approach to guarantee that CWP RH pilot case studies could easily exploit these digital resources.

5.2.2 Structure / main repositories

These repositories are structured as follows:

[4] https://www.fao.org/fi/static-media/MeetingDocuments/cwp/cwp_IS_2017/Pr2e.pdf
[5] https://www.rd-alliance.org/groups/fisheries-data-interoperability-wg
[6] https://doi.org/10.1038/sdata.2016.18

- *fdi-terms* (https://github.com/fdiwg/fdi-terms): referencing a list of fishery preferred terms, applicable for fishery concepts and their definitions, to be used as basis for data exchange, but also for dissemination (e.g., through web-application data access/queries forms)
- *fdi-codelists* (https://github.com/fdiwg/fdi-codelists): referencing all code lists (including CWP standard code lists) and their metadata. Two main sub-directories are set to distinguish between:
    - *global* code lists and
    - *regional* code lists
- *fdi-mappings* (https://github.com/fdiwg/fdi-mappings): referencing all mappings between code lists, including:
    - *national-to-regional* (mappings between national and regional code lists)
    - *national-to-global* (mappings between national and global code lists)
    - *regional-to-global* (mappings between regional and global code lists)
    - *global-to-global* (mappings between global code lists)

  Mappings are not only aimed to map one geographic level to another, but may also be applied across concepts to support data qualification during data validation. Such mappings may provide a rich source of knowledge and way to improve data quality (e.g., mapping between species codes and tRFMOs in context of FIRMS GTA to know which species are supposed to be reported by a given tRFMO; mapping between target species codes and fishing gears; mapping between species codes, and catch types).

- *fdi-formats* (https://github.com/fdiwg/fdi-formats): referencing digital format specifications to be used as basis for data templating, validation and submission in data collection frameworks use cases piloted through the FIRMS dcf-shiny platform (e.g., FIRMS Global Tuna Atlas data calls, WECAFC-FIRMS DCRF, RECOFI MDR).

5.2.3 Typographical convention

For dataset dimensions (i.e., terms) as used as column names; but also for naming resources (code lists, code list mappings), the recommended typographical convention is the *snake case*[7] where text parts should be lowercase separated by the underscore ("_") character.

Mixing lowercase and uppercase (such as in the *camel case*[8] convention) is not recommended.

By principle the lowercase should be preferred as part of the strict *snake case* convention, to foster maximum data portability and ease of use, although the uppercase may be tolerated as replacement *(screaming snake case)*.. Indeed, such consideration essentially depends on the Operating System (OS) considered and the IT model/system used for storing data:

[7] https://en.wikipedia.org/wiki/Snake_case
[8] https://en.wikipedia.org/wiki/Camel_case

- Data Base Management Systems (DBMS) may be lowercase oriented (e.g., PostgreSQL) or uppercase oriented (e.g., Oracle,SQLServer). On the choice of lowercase vs. uppercase, as example, MySQL/MariaDB official documentation indicates: *... it is best to adopt a consistent convention, such as always creating and referring to databases and tables using lowercase names. This convention is recommended for maximum portability and ease of use*. (MySQL 8.0 reference manual - https://dev.mysql.com/doc/refman/8.0/en/identifier-case-sensitivity.html)
- SDMX guidelines suggest a systematic uppercase convention, although it is noted that lower case is technically allowed by the standard (See SDMX guidelines for the creation and management of SDMX code lists - https://sdmx.org/wp-content/uploads/SDMX_Guidelines_for_CDCL.docx , page 5). There is no clear argument why lowercase should be avoided and the guidelines indicate only "*Even though technically allowed in the standard, it is highly recommended not to use lower case characters in order to avoid possible confusion and technical issues with upper case characters*".

5.2.4 Digital resources identification

5.2.4.1 Enforcing of persistent identifiers

5.2.4.1.1 Semantic identifiers

As general information management best practice to design **globally unique persistent identifiers (PIDs)**, and as joint capacity building effort with the work performed in the CWP TG-Geospatial to design and implement a Water Jurisdiction Areas classification, it was decided to try implementing a coding system aligned with the Unique Resource Name (URN) specification[9].
Considering digital resources may not always formally refer to CWP standards, but to candidate standards or ad hoc standards that are not part of CWP but identified as case studies cross-cutting needs; the URN chosen will refer to the ad hoc *fdi* namespace identifier (standing for *Fisheries Data Interoperability*) chosen for the CWP RH digital resources.

The below coding system overall scheme proposal consists in identifying the resources by type, including the authority:

urn:fdi:<resourceType>:<authority>:<resourceCode>

If adopted, through update lifecycles of the digital resources, this coding system can be further expanded to include the version of the resource. The four types of digital resources identified will be then identifiable through the base URN parts:

- *fdi-terms:* urn:fdi:term (Example: urn:fdi:term:fishing_fleet – More examples at https://github.com/fdiwg/fdi-terms)
- *fdi-codelists*: urn:fdi:codelist (Example: urn:fdi:codelist:wecafc:species – More examples at https://github.com/fdiwg/fdi-codelists). The work of CWP TG-Geospatial suggests another case study implementation of URN with a Water Jurisdiction Area code list identified as

[9] https://en.wikipedia.org/wiki/Uniform_Resource_Name

urn:fdi:codelist:cwp:wja expanding the coding system to *codes* with the following URN: urn:fdi:code:wja

- *fdi-mappings:* urn:fdi:mapping
- *fdi-formats*: urn:fdi:format

The need to set unique persistent identifiers is a general requirement for properly setting a knowledge base (broader concept than a catalogue) made of semantically identified resources. The Global Record of Stocks and Fisheries is an example of a case study that does need proper identification of resources to be able to build reusable semantic identifiers for stocks and fisheries. The approach proposed suggests relying on an existing information management standard (URN) to build up these identifiers.

5.2.4.1.2 Digital Object Identifiers

It is suggested to complement this semantic identification by web digital object identification following the DOI mechanism.

In order to keep archiving versions of digital resources (and their metadata), and recognizing the general need for digital resource citations of CWP digital resources through use cases, CWP TG-RH recommends to assign a Digital Object Identifier (DOI) to each digital resource, following similar practices of case studies such as the FIRMS GTA to assign DOIs through the Zenodo[10] open platform (based on DataCite[11]).

5.2.4.2 Versioning

CWP TG-RH suggests to start associating a version to each digital resource, by aligning on the versioning principle illustrated in previous cataloguing exercices made in CWP, i.e. following the scheme:

<year>.<major version>.<minor version>(.<correction>)

Examples: 2023.1.0; 2023.1.3.1

When the resources are tightly managed within a data lifecycle (eg ASFIS yearly update), the version should follow this lifecycle.

CWP TG-RH should further evaluate if versioning could be part of the semantic identification mechanism based on URN.

5.2.5 Metadata

Each digital resource should be described with a proper metadata that should be periodically checked for validity and maintained to reflect changes in the data resource.

[10] https://zenodo.org/
[11] https://datacite.org/

For simplicity, in order to fulfill minimum metadata requirements, resource metadata added so far follow the DublinCore[12] standard, with one DublinCore XML metadata file produced for each digital resource. Hereafter is an example of metadata produced for a WECAFC regional code list established as part of the WECAFC-FIRMS DCRF (available at https://github.com/fdiwg/fdi-codelists/blob/main/regional/wecafc/cl_fleet_segment.xml):

```
<atom:entry xmlns:atom=http://www.w3.org/2005/Atom
                xmlns:dc=http://purl.org/dc/elements/1.1/ xmlns:dcterms="http://purl.org/dc/terms/"
                xmlns:xlink="http://www.w3.org/1999/xlink"
                xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <atom:updated>2023-04-25T23:59:23</atom:updated>
  <dcterms:title>WECAFC Reference List of Fleet segments</dcterms:title>
  <dcterms:source>WECAFC Interim Data Collection Reference Framework (version 0.8):
https://www.fao.org/3/cc0657en/cc0657en.pdf</dcterms:source>
  <dcterms:identifier>urn:fdi:codelist:wecafc:fleet_segment:0.8</dcterms:identifier>
  <dcterms:identifier>cl_fleet_segment</dcterms:identifier>
  <dcterms:license>NONE</dcterms:license>
  <dcterms:hasVersion>0.8</dcterms:hasVersion>
  <dcterms:type>Dataset</dcterms:type>
  <dcterms:creator>WECAFC Working Group for fisheries data and statistics (FDS-
WG)</dcterms:creator>
  <dcterms:publisher>Fisheries Data Interoperability adhoc working group</dcterms:publisher>
  <dcterms:relation>urn:fdi:term:fleet_segment</dcterms:relation>
  <dcterms:subject>WECAFC</dcterms:subject>
  <dcterms:subject>WECAFIS</dcterms:subject>
  <dcterms:subject>DCRF</dcterms:subject>
  <dcterms:subject>FAO</dcterms:subject>
  <dcterms:subject>CWP</dcterms:subject>
  <dcterms:subject>FDI</dcterms:subject>
  <dcterms:subject>FAIR</dcterms:subject>
  <dcterms:subject>Fisheries</dcterms:subject>
  <dcterms:subject>Interoperability</dcterms:subject>
</atom:entry>
```

From these basic metadata, other more complex metadata formats could be handled depending on needs. The advantage of the DublinCore is that it can be considered as pivot metadata model quite easy to interoperate with other metadata standard formats.

For CWP geospatial code lists, a relationship should be established with existing ISO 19115 geographic metadata already maintained and disseminated through the FAO NFI GeoNetwork platform, as it is the case for FAO major areas or the CWP areal grid systems.

At this stage, not all digital resources are described with metadata. It is proposed that CWP TG-RH should attach metadata to all existing digital resources to ensure metadata coverage completeness.

The provision of these metadata resources tightly associated with the data resources that are described makes possible the collaborative edition/maintenance of the resources, fostering contribution by CWP

[12] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

parties and related partners. CWP TG-RH suggest to couple this process from the FAO master data management workflow, but in a *loosely* way so we may inherit relevant metadata coming from FAO metadata workflows where applicable, while maintaining others directly through the digital resource repositories where there is no rationale to have them maintained in FAO master data management tools.

Except on the general need to assign DOIs to digital resources, no reasonable recommendation can be made by CWP TG-RH regarding the publication of these digital resources (together with their metadata) through a particular cataloguing technology. Nonetheless, CWP TG-RH advises to give more weight to technologies implementing metadata and cataloguing standards (eg. DCAT).

5.2.6 Pilot case studies

As proof-of-concept, digital resources made available through the *fdiwg* ad hoc repositories were proven to foster implementation of CWP Reference harmonization in various case studies (**Table 1**). Some case studies may already align on the concepts and definitions provided by the CWP RH implementation guidelines but may not exploit the digital resources put disposal, in which case they are indicated as *planned*.

Table 1. Matrix describing the status of availability of digital resources for different case studies. * indicates the resources are planned to be developed

	fdi-terms	*fdi-codelists*	*fdi-mappings*	*fdi-formats*
FIRMS Global Tuna Atlas (GTA)	Yes	Yes	Yes	Yes
Global Record of Stocks and Fisheries (GRSF)	Yes	Yes	-	-
WECAFIS (WECAFC-FIRMS DCRF)	Yes	Yes	Yes*	Yes
RECOFI MDR PILOT	Yes*	Yes*	-	Yes*
FCWC regional database (Gulf of Guinea) - Industrial data	Yes*	Yes*	Yes*	-
FAO Calipseo platform (reporting component)	Yes	Yes	-	Yes
FAO NFISS Yearbook	-	Yes*	Yes*	-
FIRMS dcf-shiny platform	Yes	Yes	Yes*	Yes

The FIRMS dcf-shiny platform[13] is a software case study and is currently used as pilot for driving data submissions for WECAFIS as part of the WECAFC DCRF. A testing phase has been started for the FIRMS Global Tuna Atlas with the alignment of IOTC and ICCAT data submissions on the data exchange format technical specification hereby proposed as CWP standard.
In case of regional case studies, mappings will be supporting advanced data validation based on cross-concept mapping rules rather than mappings to move from regional to global level.

[13] https://github.com/fdiwg/dcf-shiny

5.3 Data exchange formats

5.3.1 Overview

Based on the exploitation of core digital resources made available and described above (terms, code lists, mappings), a fisheries data exchange formats framework can be then established to facilitate implementation on fisheries databases.

Based on the various feedbacks provided by case studies, two main types of data structures can be specified:

- a *generic* **measurement-agnostic data structure**, compliant with international standards for describing digital information and fisheries (e.g., ISO, CWP) and data exchange best practices; as preferred option for advanced data managers and machine readability
- a *simplified* **measurement-specific data structure**, accepting partly denormalized data to foster edition/provision by clerks in charge of preparing and exchanging data.

5.3.2 Format specifications

5.3.2.1 Digital format requirements

Recommended digital format

The recommended digital format is the Comma-Separated-Value (CSV) as defined in standard RFC 4180 with the following characteristics

- mime type: text/csv
- column separator: , (comma)
- decimal separator: . (dot)
- strings unquoted or quoted depending on the nature of the data. Handling codes with leading zeros (e.g., ISSCFG gear classification) may need a particular attention and should be considered as strings. To do so, quoting may be needed.

Other digital formats

Other digital formats for data exchange may be considered as long as they comply with the data structure definitions (see below). Since some digital formats can be managed by generic readers (through identification of mime types and sub types), other digital formats can also be supported, although **it is strongly advised to use a single digital format by use case basis**, e.g., CSV digital format for the FIRMS Global Tuna Atlas data exchange.

Digital formats that might be supported on a use case basis:

- MS Excel (*.xlsx* extension - mime type: *application/vnd.openxmlformats-officedocument.spreadsheetml.sheet*), also strongly not advised due to issues with file encoding, locales, and content alteration (e.g., dates, numbers). Such format should be considered with caution in contexts where the format is well established and adopted, and where changing the data exchange format would be too difficult to enforce. This format may

be envisaged in early stages of setting data collection frameworks, as communication tool to capture all aspects of the data exchange requirements.

- JSON (*.json* extension - mime type: application/json), which may constitute an alternative to CSV. Such format may be considered in context where fisheries data exchange is envisaged through web-services.

File encoding

When formatted in CSV and JSON, data files must be encoded as UTF-8.

5.3.2.2 Data dictionary

In order to put in place data structure definitions driving the fisheries data exchange format as envisaged standard, a data dictionary has been put in place to list the fisheries data terms to be used as column names, as well as their definitions for a reuse in structural dataset metadata (e.g., ISO 19110, SDMX DSD).

This data dictionary is available at https://github.com/fdiwg/fdi-terms

It should be emphasized that a fisheries data term is not tightly associated with a certain reference classification system and code list. It will be tightly associated to a concept but is loosely coupled to a set of classification systems and related code lists that can be different (e.g., use of ISSCFG code or standard abbreviation for gear; ISO 3 or M49 code list for flagstate; use of national, regional or global code list). This will essentially depend on the context The association with a particular code list will be done when implementing a specific data structure definition through a digital format specification, bound to a certain context.

5.3.2.3 Data structures

5.3.2.3.1 Generic (measurement-independent) data structure

The generic data structure is measurement-independent. In principle, such data structure could be extended for any data domain considered where:

- Domain-specific and spatio-temporal columns can be added, following standard classification systems, with column definitions aligned with *fdi-terms*. Classification systems can be enforced or extended as needed. Domain-specific and spatio-temporal dimension columns characterize the dimensions of data:
    - Who: administrative entity /political entity
    - When: temporal dimension of data
    - Where: area identifiers, geographic coordinates
    - How: fleet segment, fishing gear, fishing mode
    - What: species
- A building block of generic measurement-related columns are added, and filled based on classification systems that will vary depending on the target data domain:
    - **measurement**: the main variable measured (e.g., catch, effort, catch-at-size, fleet engagement)
    - **measurement_type**: the typology of measurement (e.g., for catch, the types of catches as derived from the Catch concepts diagram: landings, discards, etc.)

- **measurement_value**: the measured value
- **measurement_unit**: the unit of measure
- **measurement_obs**: an observation over the measure (terminology and content to be further discussed if it should be free text only, and ideally relying on a standard classification, e.g., SDMX OBS_STATUS classification)

Annex 1 details how such data structure formatting would apply to catch and effort data.

5.3.2.3.2 Simplified (measurement-dependent) data structure

In addition to the generic data structure, the present specification suggests a **denormalized data structure** that is **measurement-dependent and depends on a contract between data provider and ingestion system.** It offers a flexible way to exchange data, driven by specific needs to support the data providers. The flexibility is visible in:

- Accepting more formats in the time dimension that clerks can later edit by hand.
- Offering a denormalized and row-focused data structure definition, to foster the editing of measurement types (e.g., catch types) that cannot be filled out in a fully normalized structure where measurement by type has to be handled as separate rows.
- Columns accept extended standard classifications where needed, with the below column definitions (respecting the order of columns).

Measurement-dependent data structure definitions are designed to be interoperable with the generic data structure definition as measurement-independent and agnostic for processing the data, and eventually for storing the data in underlying systems (e.g., FIRMS Global Tuna atlas database, WECAFIS Regional database). Concretely, this means that:

- ISO standard time_start / time_end can be easily derived out of the year/period formats
- Measurements by type can be easily normalized towards the *measurement* build block (made of *measurement, measurement_type, measurement_value, measurement_unit and measurement_obs* **standardized columns).**

Business logic

The structure denormalized by measurement types relies on several principles:

- use of a typographical convention built on *measurement, measurement type* for naming the measurement type value column:

  <measurement>_<measurement_type>

  Example: catch_retained, catch_discarded, catch_nominal, effort_nominal

- for each measurement type value column, a dedicated unit column following the writing convention:

  <measurement>_<measurement_type>_unit

- a general measurement observation, in case additional observations/notes have to be provided.

  <measurement>_obs

In case observations may be needed over measurements by type, a similar scheme could be applied as follows:

<measurement>_<measurement_type>_obs

Based on this logic, simplified structures can be built to cover a particular measurement / data task, and data following such simplified structures can be easily normalized to the generic data structure.

Annex 2 provides simplified data structures for common data tasks covered in case studies such as FIRMS Global Tuna atlas and WECAFIS Regional database.

Annex 1: Generic data structure annotated for catch & effort data concepts & classification systems

	Term (column name)	Concept name	Classification system	Type	Definition	Target values
Who ***[ADMINISTRATIVE / POLITICAL ENTITY]***	*entity* (abstract concept ID, implemented differently depending on the context) Implementation cases: • **fishing_fleet** (FIRMS Global Tuna atlas) • **flagstate** (WECAFC-FIRMS) • country (FAO)	*Entity* (abstract concept name, implemented differently)	ISO 3166-1 (ISO 3) strictly, or extended to apply to specific context entities M49 for flagstates	string	Code of the *entity*. The definition will be defined differently depending on the actual *entity* concept name implementation	ISO3 codes (uppercase), expanded with ad hoc codes (e.g., Tuna Atlas fishing fleet code list) If not defined (case where the dimension is not needed, or optional) leave empty (NA)
When	**time**	Time start/end	ISO 8601	string	ISO 8601 format (made of YYYY-MM-DD extended to time/timezone if needed)	Concatenation <time_start>/<time_end> OR separate columns time_start / time_end
	time_start	Time start	ISO 8601	string	ISO 8601 format (YYYY-MM-DD extended to time/timezone if needed)	Value should be lower than or equal to time_end
	time_end	Time end	ISO 8601	string	ISO 8601 format (YYYY-MM-DD extended to time/timezone if needed)	Value should be greater than or equal to time_start

Where ***[GEOGRAPHIC AREA]***	**geographic_identifier**	Geographic identifier	context-specific: - Tuna atlas: CWP grid codes - WECAFC-FIRMS: FAO areas breakdown	string	Code identifying the georeferencing unit.	For gridded catches, use the CWP grid code
	geographic_coordinates	Geographic coordinates	Well-Known Text format	string	OGC WKT	Optional.
How ***[FLEET SEGMENT / FISHING PRACTICE]***	**fleet_segment**	Fleet segment (i.e. Vessel type x Length class)	ISSCFV for both vessel type and length class	string	Code of the fleet segment - vessel types by length classes	
	gear_type	Fishing gear	ISSCFG	string	Code of the fishing gear type	ISSCFG codes
	fishing_mode	Fishing mode	Adhoc classification system, see target values	string	Code of the fishing mode	Possible values (to be refined in collaboration with TG-Catch / TG-Effort) *NA* (n-a) - Not applicable *ALL* (All) - All fishing modes reported together *FREE* (Free) - Fishing on free school, no FAD use *ASSO* (Associated) - Fishing on FAD associated school *DIVE* (Diving) - Fishing with one or more divers
What	**species**	Species	ASFIS	string	Code of the species	ASFIS codes. NA if not applicable.

Measurement ***(Generic structure)***	**measurement**	Measurement	- Catches: 'catch' - Effort: 'effort'			
	measurement_type	Measurement type	- Catches: catch types following values: 'retained', 'discarded', 'nominal' - Effort: NA	string		For catches, this can be used to store the type of catch values. In the Tuna atlas catchtype codelist values are used. NA if not applicable (e.g., effort)
	measurement_value	Measurement value		number	Measured value	Live-weight equivalent ('t'). Case of FIRMS Tuna atlas level 0: mix with number of fishes ('no') as temporary solution until level 1 is drafted based on unit conversion factors.
	measurement_unit	Measurement unit		string	Code of the unit	Values authorized are: • for catches: "t" (metric tonnes), "no" (number of fish) ad hoc unit (in principle = no unit) • for effort: Tuna atlas use of Tuna atlas effort codelist ?
	measurement_obs	Measurement observation		string	Value observations	Best practice in statistical series handling, required in case notes have to be set for specific series values (e.g., estimation/assumption made)

Annex 2: Simplified data structures annotated for catch and effort, and fleet engagement (case of WECAFC DCRF task III.1)

Catch measurement

	Term (column name)	Concept name	Classification system	Type	Definition	Target values
Who	*entity* (abstract concept name, implemented differently depending on the context) Implementation cases: • **fishing_fleet** (FIRMS Global Tuna atlas) • **flagstate** (WECAFC-FIRMS/RECOFI)	*Entity* (abstract concept name, implemented differently)	ISO 3166-1 (ISO 3) strictly, or extended to apply to specific context entities M49 for flagstates	string	Code of the *entity*. The definition will be defined differently depending on the actual *entity* concept name implementation	ISO3 codes (uppercase), extended with ad hoc codes eventually (e.g., FIRMS Tuna atlas code list) If not defined in strata use "ALL"
When	**year**	Year		integer	Year	
	period	Period	Ad hoc	string	Quarter / month definition	Non standard - for months: M1 to M12 - for quarters: Q1 to Q4
Where	**geographic_identifier**	Geographic identifier	context-specific: - Tuna atlas: CWP grid codes - WECAFC-FIRMS/RECOFI: FAO areas breakdown	string	Code identifying the georeferencing unit.	For gridded catches, use the CWP grid code

	geographic_coordinates	Geographic coordinates	Well-Known Text format	string	OGC WKT	Optional
How	**fleet_segment**	Fleet segment / Vessel type	ISSCFV	string	Code of the fleet segment - vessel types by length classes	
	gear_type	Gear type	ISSCFG	string	Code of the fishing gear type	ISSCFG codes
	fishing_mode	Fishing mode		string	Code of the fishing mode	Possible values *NA* (n-a) - Not applicable *ALL* (All) - All fishing modes reported together *FREE* (Free) - Fishing on free school, no FAD use *ASSO* (Associated) - Fishing on FAD associated school *DIVE* (Diving) - Fishing with one or more divers
What	**species**	Species	ASFIS	string	Code of the species	ASFIS codes. NA if not applicable
Catch measurement	**catch_retained**	Retained catches		number	Retained catch measurement value	
	catch_retained_unit	Retained catches unit		string	Code of the unit	Values authorized are: for catches: “t” (metric tonnes), “no” (number of fishes) ad hoc unit (in principle = no unit)
	catch_discarded	Discarded catches		number	Discarded catch measurement value	
	catch_discarded_unit	Discarded catches unit		string	Code of the unit	Values authorized are: for catches: “t” (metric tonnes), “no” (number of fishes) ad hoc unit (in principle = no unit)

	catch_nominal	Nominal catches		number	Nominal catch measurement value	
	catch_nominal_unit	Catch measurement unit		string	Code of the unit	Values authorized are: for catches: “t” (metric tonnes), “no” (number of fishes) ad hoc unit (in principle = no unit)
	catch_obs	Catch measurement observation		string	Value observations	

Effort measurement

	Term (column name)	Concept name	Classification system	Type	Definition	Target values
Who	*entity* (abstract concept name, implemented differently depending on the context) Implementation cases: • **fishing_fleet** (FIRMS Global Tuna atlas) • **flagstate** (WECAFC-FIRMS)	*Entity* (abstract concept name, implemented differently)	ISO 3166-1 (ISO 3) strictly, or extended to apply to specific context entities M49 for flagstates	string	Code of the *entity*. The definition will be defined differently depending on the actual *entity* concept name implementation	ISO3 codes (uppercase), extended with ad hoc codes eventually (e.g., FIRMS Tuna atlas code list) If not defined in strata use "ALL"
When	**year**	Year		integer	Year	
	period	Period	Ad hoc	string	Quarter / month definition	Non standard - for months: M1 to M12 - for quarters: Q1 to Q4
Where	**geographic_identifier**	Geographic identifier	context-specific:	string	Code identifying the	For gridded catches, use the CWP grid code

			- Tuna atlas: CWP grid codes - WECAFC-FIRMS: FAO areas breakdown		georeferencing unit.	
	geographic_coordinates	Geographic coordinates	Well-Known Text format	string	OGC WKT	Optional
How	**fleet_segment**	Fleet segment / Vessel type	ISSCFV	string	Code of the fleet segment - vessel types by length classes	
	gear_type	Gear type	ISSCFG	string	Code of the fishing gear type	ISSCFG codes
	fishing_mode	Fishing mode		string	Code of the fishing mode	Possible values *NA* (n-a) - Not applicable *ALL* (All) - All fishing modes reported together *FREE* (Free) - Fishing on free school, no FAD use *ASSO* (Associated) - Fishing on FAD associated school

						DIVE (Diving) - Fishing with one or more divers
What	**species**	Species	ASFIS	string	Code of the species	ASFIS codes. NA if not applicable.
Effort measurement	**effort_fishing_duration**	Fishing duration		number		e.g., WECAFC-FIRMS Nominal effort → days fishing
	effort_fishing_duration_unit	Fishing duration unit		string	Code of the fishing duration unit	e.g., WECAFC-FIRMS Nominal effort → days
	effort_nominal	Nominal Effort		number	Effort measurement value	
	effort_nominal_unit	Nominal Effort measurement unit		string	Code of the nominal effort unit	Values authorized are: hooks, fishing days (all fisheries, number of fishing operations
	effort_obs	Effort measurement observation		string	Value observations	Free text for now

Fleet engagement

	Term (column name)	Concept name	Classification system	Type	Definition	Target values
Who	*entity* (abstract concept name, implemented differently depending on the context) Implementation cases: • **fishing_fleet** (FIRMS Global Tuna atlas) • **flagstate** (WECAFC-FIRMS)	*Entity* (abstract concept name, implemented differently)	ISO 3166-1 (ISO 3) strictly, or extended to apply to specific context entities M49 for flagstates	string	Code of the *entity*. The definition will be defined differently depending on the actual *entity* concept name implementation	ISO3 codes (uppercase), extended with ad hoc codes eventually (e.g., FIRMS Tuna atlas codelist) If not defined in strata use "ALL"
When	**year**	Year		integer	Year	
	period	Period	Ad hoc	string	Quarter / month definition	Non standard - for months: M1 to M12 - for quarters: Q1 to Q4
Where	**geographic_identifier**	Geographic identifier	context-specific: - Tuna atlas: CWP grid codes	string	Code identifying the georeferencing unit.	For gridded catches, use the CWP grid code

			- WECAFC-FIRMS: FAO areas breakdown			
	geographic_coordinates	Geographic coordinates	Well-Known Text format	string	OGC WKT	Optional
How	**fleet_segment**	Fleet segment / Vessel type	ISSCFV	string	Code of the fleet segment - vessel types by length classes	
	gear_type	Gear type	ISSCFG	string	Code of the fishing gear type	ISSCFG codes
What	**species**	Species	ASFIS	string	Code of the species	ASFIS codes. NA if not applicable.
Effort measurement	**fleet_engagement_number**	Fleet engagement number		number		e.g., number of vessels
	fleet_engagement_number_unit	Fleet engagement number unit (number of vessels)		string		N/A. Optional column

fleet_engagement_capacity	Total capacity		number	Effort measurement value	
fleet_engagement_capacity_unit	Total capacity unit		string	Code of the fleet engagement capacity unit	
fleet_engagement_enginepower	Total engine power		number		
fleet_engagement_enginepower_unit	Total engine power unit		string	Code of the fleet engagement engine power unit	
fleet_engagement_obs	Effort measurement observation		string	Value observations	Free text for now